HHS Public Access

Author manuscript

Mol Carcinog. Author manuscript; available in PMC 2018 March 01.

Published in final edited form as:

Mol Carcinog. 2017 March ; 56(3): 1000–1009. doi:10.1002/mc.22565.

Single nucleotide variants in metastasis-related genes are associated with breast cancer risk, by lymph node involvement and estrogen receptor status, in women with European and African ancestry

Michelle R. Roberts[1,2,3], **Lara E. Sucheston-Campbell**[4], **Gary R. Zirpoli**[5], **Michael Higgins**[6], **Jo L. Freudenheim**[3], **Elisa V. Bandera**[7], **Christine B. Ambrosone**[2], and **Song Yao**[2]

[1]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[2]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY

[3]Department of Epidemiology and Environmental Health, University at Buffalo, Buffalo, NY

[4]Division of Pharmacy Practice and Science, The Ohio State University, Columbus, OH

[5]Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA

[6]Department of Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo, NY

[7]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ

Abstract

Background—Single nucleotide polymorphisms (SNPs) in pathways influencing lymph node (LN) metastasis and estrogen receptor (ER) status in breast cancer may partially explain inter-patient variability in prognosis. We examined 154 SNPs in 12 metastasis-related genes for associations with breast cancer risk, stratified by LN and ER status, in European-American (EA) and African-American (AA) women.

Methods—2,671 women enrolled in the Women's Circle of Health Study were genotyped. Pathway analyses were conducted using the adaptive rank truncated product (ARTP) method, with $p_{ARTP} \leq 0.10$ as significant. Multi-allelic risk scores were created for the ARTP-significant gene(s). Single-SNP and risk score associations were modeled using logistic regression, with false discovery rate (FDR) p-value adjustment.

Results—Although single-SNP associations were not significant at $p_{FDR} < 0.05$, several genes were significant in the ARTP analyses. In AA women, significant ARTP gene-level associations included *CDH1* with LN+ ($p_{ARTP}=0.10$; multi-allelic OR=1.13, 95% CI 1.07–1.19, $p_{FDR}=0.0003$) and *SIPA1* with ER− breast cancer ($p_{ARTP}=0.10$; multi-allelic OR=1.16, 95% CI 1.02–1.31, $p_{FDR}=0.03$). In EA women, *MTA2* was associated with overall breast cancer risk ($p_{ARTP}=0.004$), regardless of ER status, and with LN− disease ($p_{ARTP}=0.01$). Also significant were *SATB1* in ER−

Corresponding author: Michelle R. Roberts, Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, michelle.roberts@channing.harvard.edu.

(p_{ARTP}=0.03; multi-allelic OR=1.12, 95% CI 1.05–1.20, p_{FDR}=0.003) and *KISS1* in LN− (p_{ARTP}=0.10; multi-allelic OR=1.18, 95% CI 1.08–1.29, p_{FDR}=0.002) analyses. Among LN+ cases, significant ARTP associations were observed for *SNAI1*, *CD82*, *NME1*, and *CTNNB1* (multi-allelic OR=1.09, 95% CI 1.04–1.14, p_{FDR}=0.001).

Conclusion—Our findings suggest that variants in several metastasis genes may affect breast cancer risk by LN or ER status, although verification in larger studies is required.

Keywords

Breast neoplasms; Single nucleotide polymorphism; Lymph nodes; Estrogen receptor; African-American

1. Introduction

Breast cancer prognosis is excellent when diagnosed and treated at an early stage, but is poorer when metastatic disease is present. The presence of axillary lymph node (LN) metastases at diagnosis is a strong risk factor for future recurrence and poorer prognosis. Once metastasized to distant sites, breast cancer is generally considered incurable and the majority of breast cancer-associated mortality results from metastatic disease [1]. Breast cancer mortality rates are higher among African-American (AA) women compared to European-American (EA) women, for reasons that are not fully understood. AA women tend to present with breast cancer at a younger age, with more aggressive tumor characteristics and a greater likelihood of disease progression and recurrence [2,3]. It is therefore critical to identify mechanisms of metastasis, particularly those that may differentially affect AA and EA patients.

Inherited genetic variation in metastasis-associated genes might partially explain inter-patient variability in successful metastatic dissemination and colonization [4]. A great number of metastasis-associated genes have been identified, including those involved in epithelial-mesenchymal transition (EMT), metastasis suppressor genes, and others [5].

EMT has been hypothesized as a mechanism by which tumor cells acquire metastatic potential, and genes involved in this process include E-cadherin (*CDH1*), Snail (*SNAI1*), and β-catenin (*CTNNB1*) [6]. SATB homeobox 1 (*SATB1*) is a genome organizing protein, which has been shown to affect expression of many genes involved in metastasis, including EMT-related genes [7]. Metastasis suppressor genes are those that can inhibit metastatic formation without affecting primary tumor growth, and include *BRMS1*, *CDH1*, *CD82/KAI1*, *KISS1*, and *NME1* [8]. These genes belong to diverse pathways, including gene transcription, cell adhesion, extracellular matrix remodeling, and apoptosis. Metastasis-associated 1 family members (*MTA1*, *MTA2*, and *MTA3*) are part of estrogen receptor (ER) signaling pathways and also interact with the EMT-related genes *SNAI1* and *CDH1* [9]. Finally, the metastasis efficiency modifying gene *SIPA1* alters cell adhesion [10] and promotes metastasis *in vivo* [11].

The purpose of this study was to examine common genetic variants in metastasis-related genes for associations with breast cancer risk, stratified by ER and LN status, and likelihood

of ER− and LN+ tumors in both EA and AA women, using a combination of single-SNP and gene-based analyses. We selected 12 metastasis-associated genes (*BRMS1*, *CDH1*, *CD82/KAI1*, *CTNNB1*, *KISS1*, *MTA1*, *MTA2*, *MTA3*, *NME1*, *SATB1*, *SIPA1*, and *SNAI1*) for analysis, focusing on those with few or no previously published studies in breast cancer.

2. Materials and Methods

2.1. Study population

The Women's Circle of Health Study (WCHS) is an ongoing case-control study designed to examine the role of genetic and non-genetic factors in relation to risk of breast cancer in AA and EA women. The study design, enrollment criteria, and collection of biospecimens and questionnaire data have previously been described in detail [12,13]. Eligible cases were women who self-identified as AA or EA, were 20–75 years of age at diagnosis, had no previous history of cancer other than non-melanoma skin cancer, were diagnosed with primary, incident, histologically confirmed invasive breast cancer or ductal carcinoma in situ (DCIS), and were English speaking. In New York City, cases were identified through collaborating hospitals in Manhattan, Brooklyn, Bronx, and Queens, and in New Jersey, rapid case ascertainment using the State Cancer Registry was conducted. Controls were identified contemporaneously using random digit dialing and had the same inclusion criteria as cases, but with no history of any cancer diagnosis other than non-melanoma skin cancer. Controls were frequency matched to cases by self-reported race, 5-year age categories, and telephone exchange (New York City) or county of residence (New Jersey). In New Jersey, AA controls were also invited to participate through community recruitment events [14]. Following agreement to participate, in-person interviews were conducted to complete informed consent and an extensive epidemiologic questionnaire. Blood and/or saliva samples were collected for later extraction of DNA. Tumor characteristics were abstracted from pathology reports.

This study was approved by the Institutional Review Boards at Roswell Park Cancer Institute (RPCI), the Rutgers Cancer Institute of New Jersey (CINJ), the Icahn School of Medicine at Mount Sinai, and the participating hospitals in New York City.

2.2. DNA sample preparation

Blood and saliva were collected as sources of genomic DNA, which was isolated from blood using FlexiGene™ DNA isolation kits (Qiagen Inc., Valencia, CA) and from Oragene™ (DNA Genotek Inc., Kanata, Ontario, Canada) saliva sample collection kits, according to the respective manufacturer's protocols. Genomic DNA was evaluated and quantitated by Nanodrop UV-spectrometer (Thermo Fisher Scientific Inc., Wilmington, DE) and PicoGreen-based fluorometric assay (Molecular Probes, Invitrogen Inc., Carlsbad, CA), and stored at −80°C until analysis.

2.3. SNP selection

SNPs for the studied genes were chosen using the SNPInfo candidate gene SNP selection pipeline (National Institutes of Environmental Health Sciences (http://snpinfo.niehs.nih.gov/)) [15], which selects multi-population tag SNPs based on HapMap

genotype and linkage disequilibrium (LD) data. For each gene, SNPs were identified for the CEU (Utah residents with Northern and Western European ancestry) and YRI (Yoruba in Ibadan, Nigeria) populations using the following settings: 5kb upstream and downstream flanking regions, 0.8 tagging proportion cutoff, 0.05 minor allele frequency (MAF) cutoff, 0.8 LD threshold, minimum of two SNPs tagged by a given tag SNP, and minimum of one tag SNP tagged per gene. In addition to the tagging SNPs, we also identified several validated SNPs using the National Center for Biotechnology Information dbSNP database (http://www.ncbi.nlm.nih.gov/snp/) that had MAF >0.05 in the CEU and/or YRI populations and were either located in coding or near gene/untranslated regions or had been previously studied (rs1052566, rs3116068 (*BRMS1*); rs2306364, rs3741378, rs75894763 (*SIPA1*)). A total of 154 tag and candidate SNPs were ultimately selected for genotyping (Supplementary Table 1).

2.4. Genotyping and quality control

Genotype and quality control/assurance methods have been described in detail previously [13]. Genotyping was conducted on all participants who were enrolled in the study through June 2011 and had sufficient DNA available for analysis. SNPs were genotyped using the Illumina GoldenGate assay (Illumina Inc., San Diego, CA), with five percent duplicates and two sets of in-house trio samples included for quality control purposes. The concordance among blind duplicate pairs was >99.9%. The average successful genotyping rate for each sample and each SNP was >95%. As shown in Supplementary Table 1, 22 SNPs were excluded from further analyses because they were monomorphic (n=1), violated Hardy-Weinberg equilibrium ($p<0.00001$; n=8), or had low call rate (<98%) in EA and/or AA controls (n=9), or had MAF <5% in both EA and AA controls (n=4), leaving 132 SNPs for analysis.

2.5. Genetic ancestry estimate

A previously validated panel of 100 ancestry informative markers (AIMs) was also genotyped to ascertain genetic ancestry and control for population admixture [16]. Ninety-five AIMs were successfully genotyped and, based on this genotype data, estimates of European and African ancestry were obtained using the STRUCTURE program [17]. We excluded 41 women with self-reported race as “other” and 12 women with ≥85% estimated ancestry discordant with their self-reported race (n=11 in AAs and n=1 in EAs). Proportion of EA genetic ancestry was included as a continuous covariate in statistical analyses.

2.6. Statistical analysis

The final dataset included 2,671 women: 658 EA cases, 649 EA controls, 621 AA cases, and 743 AA controls. ER and LN status was available for 943 and 975 cases, respectively. All analyses were conducted for EA and AA participants separately. Demographic variables and tumor characteristics were compared between groups using Chi-squared, Fisher’s exact, or t-tests, as appropriate. Genotype and allele frequencies were compared between EA and AA controls using Chi-squared tests, and a Bonferroni correction was applied to adjust for multiple comparisons.

Unconditional logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (CI) for 132 SNPs in association with the study outcomes (risk of breast cancer in all cases, risk of breast cancer limited to invasive cases, risk of breast cancer stratified by LN and ER status, and case-case comparisons of LN and ER status). For each SNP, codominant, dominant, recessive, and additive genetic models were tested, using the homozygous common genotype in the EA group as the reference category. Age and proportion of EA genetic ancestry were chosen *a priori* as covariates for single-SNP analyses. All analyses were performed separately in AA and EA women, using the same genotype as reference category to facilitate comparisons between the two groups. False discovery rate testing was implemented to control for multiple comparisons [18]. Chi-squared and t-tests, logistic regression, and false discovery rate analyses were performed using SAS version 9.3.

To consider all markers jointly, pathway analyses were conducted using the adaptive rank truncated product method (ARTP). The ARTP method is a gene-based approach to pathway analysis that can confer a statistical power advantage in situations in which the causal variant is located in genes with fewer SNPs within a pathway, because it limits the effect of many SNPs in the larger gene(s) with null associations [19]. SNPs with MAF <0.05 in either EA (n=22) or AA (n=1) control groups were excluded from ARTP analysis in that group, leaving a total of 110 and 131 SNPs available for analysis in EA and AA participants, respectively. LD statistics generated for all SNP pairs in EA and AA controls were used to further filter SNPs in strong LD, defined as r^2 >0.80 between any two markers. For each pair or group of SNPs in strong LD, the SNP with the strongest single-SNP association was retained. This resulted in a final set of 87 independent SNPs for EA participants and 122 independent SNPs for AA participants. For power considerations, SNPs with variant homozygous genotype frequencies ≤10% were collapsed to combine the heterozygous and variant homozygous genotype categories.

The ARTP method relies on user-defined pathways. To identify interaction networks between the 12 genes we selected for genotyping, we employed the GeneMANIA gene interaction prediction server (www.genemania.org) [20], using the default weighting method. We considered those genes that interacted with each other to form a pathway. The ARTP method was then used to obtain gene and pathway p-values for associations with each outcome. Permutation p-values were obtained by permuting the outcome vector (i.e., case-control status, ER status, or LN status, as appropriate) 10,000 times, while adjusting for age and proportion of European ancestry, using the default truncation points. Since permutation testing is a very conservative method for controlling the family-wise Type I error rate, we chose the more liberal p ≤0.10 as the gene and/or pathway significance threshold. Analyses were performed using the R package “ARTP”, version 2.0.4 (R Foundation for Statistical Computing, Vienna, Austria).

The ARTP method provides an overall test of whether variants in each gene or pathway are associated with risk, but it does not provide an estimate of the magnitude or direction of the association. Therefore, multi-allelic risk scores were constructed to estimate the risk of breast cancer associated with genes found to be significant in ARTP analyses [21]. For each outcome in which one or more genes were significant at p_{ARTP}≤0.10, SNPs in these significant genes were included if the additive single-SNP model OR was <0.95 or >1.05. If

the additive model OR for a given SNP was between 0.95–1.05, the association was considered null and the SNP was excluded from the summary score. For each SNP, the direction of the association with its respective outcome (positive or inverse) was determined using the additive single-SNP model. SNPs with positive associations were assigned 0, 1, and 2 at-risk alleles, and SNPs with inverse associations were assigned 2, 1, and 0 at-risk alleles. For each outcome, the summary score SNPs were added together; participants missing genotype for one or more SNPs were not assigned a summary score, to avoid misclassification with respect to number of at-risk alleles.

Summary scores were included as continuous variables in logistic regression models. Age and proportion of European ancestry were included as covariates in minimally adjusted models. Participant and tumor characteristics that were significant at $p<0.05$ in EA or AA comparisons were included as additional covariates in fully adjusted models. P-values derived from the logistic regression models were adjusted for multiple comparisons using the FDR method.

3. Results

Selected participant characteristics are shown in Table 1. When the case group was restricted to those with invasive tumors, the distribution of characteristics between cases and controls was similar among both EA and AA participants (data not shown). When compared by ER and LN status, characteristics were generally similar for both EA and AA cases. LN− cases tended to be older than LN+ cases in both ancestral groups. Among AA cases, women who were LN+ were significantly more likely to be premenopausal than those who were LN− (data not shown).

Genotype and allele frequencies were compared between EA and AA controls and results are shown in Supplementary Table 2. For most of the SNPs we measured, genotype and allele frequencies were significantly different between the groups, after Bonferroni correction for multiple comparisons.

Results of single-SNP analyses, under the additive and dominant models, for comparisons by ER status and LN status are shown in Supplementary Tables 3 and 4, respectively. After FDR adjustment, none of the associations remained significant.

Using GeneMANIA, we found that, by including the *ESR1* (ERα) gene, all of the metastasis-associated genes selected for this study could be linked, as shown in Supplementary Figure 1. We therefore considered the 12 genes included in this study to be components of a “metastasis pathway” for ARTP analysis. The $\log_{10}$-transformed gene and pathway p-values for each outcome are shown in Figure 1; the raw p-values are provided in Supplementary Table 5.

3.1. ARTP results in EA women

In EA women (Figure 1A), the overall pathway was significant only when all cases and controls were compared ($p_{ARTP}=0.10$). In the analysis of all cases vs. controls and when restricted to cases with invasive tumors vs. controls, the *MTA2* gene was highly significant

(p_{ARTP}=0.004 and p_{ARTP}=0.01, respectively). Risk of LN+ disease was significantly associated with *SNAI1* (p_{ARTP}=0.10), *CD82* (p_{ARTP}=0.05), *NME1* (p_{ARTP}=0.10), and *CTNNB1* (p_{ARTP}=0.09). When LN− cases were compared to controls, the *KISS1* (p_{ARTP}=0.10) and *MTA2* (p_{ARTP}=0.01) genes were significant. In case-case analysis, positive LN status was significantly associated with the *SNAI1* (p_{ARTP}=0.01) and *CTNNB1* (p_{ARTP}=0.06) genes. When stratified by ER status, the *MTA2* gene (p_{ARTP}=0.02) was significantly associated with risk of ER+ breast cancer, while the *MTA2* (p_{ARTP}=0.08) and *SATB1* (p_{ARTP}=0.03) genes were significantly associated with risk of ER− breast cancer. None of the genes were significant in case-case analysis.

3.2. ARTP results in AA women

In AA women (Figure 1B), fewer significant associations were observed. Risk of LN+ breast cancer was significantly associated with the *CDH1* gene (p_{ARTP}=0.10). In case-case analysis, *SIPA1* was significantly associated with negative ER status (p_{ARTP}=0.09) and when compared to controls (p_{ARTP}=0.10). None of the overall pathways were significant, nor were genes in any of the other outcomes.

3.3. Multi-allelic risk score results in EA women

In EA women, risk scores made from two SNPs in *MTA2* were associated with increased risk of breast cancer among all cases and when limited to invasive breast cancer in fully adjusted models (Table 2; all cases vs. controls OR=1.15, 95% CI 1.04–1.26, p_{FDR}=0.01; invasive cases vs. controls OR=1.19, 95% CI 1.07–1.33, p_{FDR}=0.003). When combined, twenty-seven SNPs in the *SNAI1*, *CD82*, *NME1*, and *CTNNB1* genes were significantly associated with risk of LN+ breast cancer (OR=1.09, 95% CI 1.04–1.14, p_{FDR}=0.001). Similar findings were observed for the comparison of LN− cases to controls, using a risk score composed of five SNPs in the *MTA2* and *KISS1* genes (OR=1.18, 95% CI 1.08–1.29, p_{FDR}=0.002). Two SNPs in *MTA2* and 13 SNPs in *MTA2* and *SATB1* were used to construct risk scores for ER+ and ER− cases, respectively, compared to controls. The *MTA2*-score was significantly associated with risk of ER+ breast cancer (OR=1.16, 95% CI 1.05–1.30, p_{FDR}=0.01), and the *MTA2/SATB1*-score with ER− disease (OR=1.12, 95% CI 1.05–1.20, p_{FDR}=0.003). In case-case analysis, the risk score comprised of six SNPs in *SNAI1* and *CTNNB1* was significantly associated with positive LN status, even when adjusting for HER2 status, tumor grade, and tumor size (OR=1.15, 95% CI 1.01–1.31, p_{FDR}=0.04).

3.4. Multi-allelic risk score results in AA women

In AA women, the multi-allelic risk score composed of 19 SNPs in the *CDH1* gene was significantly associated with increased risk of LN+ breast cancer (Table 2; OR=1.13, 95% CI 1.07–1.19, p_{FDR}=0.0003). Five SNPs in *SIPA1* were used to make risk scores for analyses of ER status in case-control and case-case analyses. When compared to controls, the *SIPA1* risk score was significantly associated with increased risk of ER− breast cancer (OR=1.16, 95% CI 1.02–1.31, p_{FDR}=0.03). In case-case analysis, however, the association was not significant when adjusted for PR status, tumor grade, tumor size, and stage.

4. Discussion

In this study, although no single-SNP association remained significant after correction for multiple comparisons, the gene-based ARTP analyses, which are implicitly controlled for multiple comparisons via permutation testing, revealed several significant gene-level associations in EA (*MTA2* with overall, ER+, ER−, and LN− breast cancer; *SATB1* with ER − breast cancer; *KISS1* with LN− breast cancer; and *SNAI1*, *CTNNB1*, *CD82*, and *NME1* with LN+ breast cancer) and AA (*CDH1* with LN+ breast cancer; *SIPA1* with ER− breast cancer) participants.

MTA2 has histone deacetylase activity and is a subunit of Mi-2/NuRD chromatin remodeling complexes [22]. By interacting with *Twist1*, a transcription factor that is a master regulator of EMT, *MTA2* mediates repression of E-cadherin [23]. Furthermore, by binding to ERα and acting as a repressor, *MTA2* modulates acetylation and transcriptional activity of ERα [24]. One of the two SNPs included in the gene-based analysis, rs11231156, is predicted by RegulomeDB [25] to be likely to affect transcription binding and is linked to expression of a gene target (RegulomeDB score 1f).

SATB1 is a nuclear matrix and scaffold attachment region binding protein [26] that acts as a genome organizer [27]. Gene expression profiling of the metastatic breast cancer cell line MDA-MB-231 has shown that, among others, *SATB1* downregulates *CTNNB1*, *CDH1*, *BRMS1*, *CD82*, *KISS1*, and *NME1* [7]. One study examining SNPs in *SATB1* with respect to breast cancer found that homozygous carriers of the −3600T/−3363A/−2984C haplotype had improved overall survival [28]. We genotyped rs1475469, which is in strong LD with −2984C>T (rs6762753; r^2=0.96; CEU population, 1000 Genomes Phase 3 data). Furthermore, using the GRASP (Genome-wide Repository of Associations between SNPs and Phenotypes) database [29], we found that two SNPs included in the gene-based analysis, rs4129096 and rs9714119, were associated with breast cancer mortality at p=0.01 and p=0.0006, respectively [30].

The transcription factor *SNAI1* is a key developmental EMT regulator and transcriptional repressor of E-cadherin and ERα [31–33]. Two of the variants in the gene-based analysis, rs6020177 and rs6091080, are predicted by RegulomeDB to affect binding (scores of 1f and 2b, respectively). While the role of *SNAI1* variants in LN metastasis has not previously been examined, the variant allele of the nonsynonymous SNP rs4647958 has been associated with reduced overall risk of breast cancer among women with older age at first pregnancy [34].

The transcriptional co-factor and structural protein β-catenin (*CTNNB1*) links E-cadherin (*CDH1*) to the actin cytoskeleton and is the central protein of the Wnt signaling pathway, which regulates cell proliferation, differentiation, and apoptosis [35]. Two variants (rs4135385 and rs1307263) have previously been examined using a candidate gene approach, although not in EA or AA populations, with conflicting results as to breast cancer risk [36,37].

CD82, *NME1*, and *KISS1* are metastasis suppressor genes, which inhibit metastatic formation without affecting primary tumor growth. *CD82/KAI1* is a cell surface glycoprotein that interacts with the Duffy antigen chemokine receptor, an endothelial cell

surface protein, to anchor tumor cells to vascular endothelial cells, inducing tumor cell senescence and suppressing metastatic spread [38].

NME1 (also known as NM23-H1) is regulated by p53 and estrogen [39,40], and affects expression of genes involved in cell migration, apoptosis, and angiogenesis through its interaction with ERα [41]. Two promoter region SNPs (rs2302254 and rs16949649), both included in our analysis, have been implicated in an increased risk of relapse, metastasis, and breast cancer-specific mortality [42] and are predicted by RegulomeDB to be likely to affect binding (scores of 2b and 1f, respectively).

KISS1 appears to be involved in the NF-κB and ER signaling pathways. Induction of *KISS1* expression resulted in decreased expression of Snail2, a transcription factor involved in EMT, and an increase in E-cadherin expression [43], indicating that one metastasis suppressive function of *KISS1* is to inhibit EMT and maintain the epithelial phenotype. One SNP we examined, rs3795573, was predicted to affect binding (RegulomeDB score 2b). One prior study found that rs5780218 was significantly associated with risk of breast cancer among Mexican patients [44], but this SNP is not in LD with any of the ones we included.

CDH1 is a tumor suppressor and metastasis suppressor, the loss of which promotes tumor cell invasiveness and induces EMT in experimental models. E-cadherin (*CDH1*) is a calcium-dependent cell-cell adhesion protein that is crucial for maintaining cell polarity, epithelial architecture and structural integrity, and preserving cell-cell interactions [35]. Several studies have examined polymorphisms in *CDH1* in relation to breast cancer risk or prognosis, with conflicting results [34,45–50], although few data exist for AA populations. Four of the SNPs included in the gene-based analysis are predicted to affect binding by RegulomeDB (rs7188750, 1b; rs9941051, 1f; rs10431923, 1f; rs9940250, 2b).

SIPA1 catalyzes the hydrolysis of guanosine triphosphate to guanosine diphosphate and has been shown to regulate cell adhesion [10]. Several *SIPA1* variants have been studied, mainly in white populations, but associations with breast cancer risk and survival have not been consistent [51,52]. In one study, rs931127 (promoter region) and rs746429, a synonymous SNP (Ala920Ala), were associated with increased likelihood of having nodal metastases at diagnosis, while a nonsynonymous SNP, rs3741378 (Ser182Phe), was associated with ER and PR negative tumors [53]. We previously observed an increased likelihood of the HER2-expressing subtype, which includes ER− tumors, in women with the variant rs3741378 genotype, although the estimate was imprecise due to small numbers [54]. Here, we found that *SIPA1* variants, including rs3741378, were associated with ER− breast cancer, in case-control and case-case analyses, in AA women.

Strengths of this study include the use of in-person interviews to gather detailed information on family history and hormone-related variables, which permitted greater control of potential confounding variables. In addition, this study used a panel of 100 AIMs to address population stratification within categorical racial groups, a well-known source of confounding due to admixture in genetic epidemiology studies. AIMs allow for estimation of individual ancestral proportions, which can then be included as a covariate in multivariate models. African populations are known to be more genetically diverse than non-African

populations [55]. Admixture with individuals of European ancestry has further contributed to genetic diversity in AAs. The panel we selected has been shown to reliably estimate the proportion of European ancestry in AA individuals, using data from the Black Women's Health Study [16]. Since these AIMs were designed to capture admixture in AAs, it is unclear how well they capture admixture in EAs; however, since there is less admixture in EA groups, confounding is less likely to have affected our results.

Our study is limited, however, by the inclusion of only 12 genes; it is likely that other metastasis-related genes contribute to breast cancer development, particularly LN+ tumors, perhaps by interacting with the genes we have investigated here. We examined common variants by identifying a set of tag SNPs within each gene, but there may be other important genetic contributions that we were unable to examine because of sample size limitations, such as rare variation and gene-environment interactions. We also restricted our definition of gene region to +/− 5kb upstream and downstream. However, since SNPs that affect gene expression may be found at distances further than this, it is possible that SNPs with important functional effects were missed by our definition.

We did not have recurrence or survival data available to examine long-term outcomes, which may be important for understanding the effects of these variants on metastasis, and our sample size precluded stratification by tumor subtype. Furthermore, we were unable to compare early-stage and metastatic breast cancer cases, due to sample size limitations; our study contained only one EA and four AA stage IV cases. Instead, we examined associations with LN status, a strong prognostic variable, and ER status, since ER− tumors are typically more aggressive and have a worse prognosis. Although ER status and LN status were missing for 26% and 12% of all cases and invasive tumor cases, respectively, participant characteristics were generally similar between those with and without known ER and LN status, indicating that our estimates were unlikely to be biased by missingness for these features.

Our study is also limited by the lack of validation using an independent population, as well as the fact that the sample sizes within subgroups prohibit separation into testing and training sets for cross-validation strategies. Finally, given the small numbers of participants in our subgroup analyses, our study may be underpowered to detect small effect sizes for the genes we examined. We therefore cannot rule out the possibility of false negative findings. For these reasons, additional large, well-powered studies with long-term outcome data are needed both to replicate the findings we report herein as well as further investigate the contribution of genetic variation to metastasis.

In conclusion, we found evidence suggesting that variants in different metastasis-related genes may affect risk of breast cancer, by LN and ER status, in EA and AA women. Several promising associations were identified that require confirmation and, in particular, investigation of SNP effects on recurrence and survival. Additional studies are needed to better understand the genetic basis for the development of breast cancer in different ancestral groups, particularly as it relates to aggressive tumor subtypes and metastatic potential.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding acknowledgement:

This work was supported in part by grants from the US National Institutes of Health (R01 CA100598, P01CA151135, and P30CA072720), US Army Medical Research and Material Command (DAMD-17-01-1-0334), the Breast Cancer Research Foundation, and a gift from the Philip L. Hubbell family. MRR was supported by DAMD W81XWH-11-1-0024, a predoctoral fellowship from the Department of Defense Breast Cancer Research Program, and R25-CA11395102, an NCI predoctoral award. The New Jersey State Cancer Registry is supported by the National Program of Cancer Registries of the Centers for Disease Control and Prevention under cooperative agreement 1US58DP003931-01 awarded to the New Jersey Department of Health. The collection of New Jersey cancer incidence data is also supported by the Surveillance, Epidemiology, and End Results program of the National Cancer Institute under contract N01-PC-2010-0027 and the State of New Jersey. The funding agents played no role in the design, collection, analysis, and interpretation of data, in the writing of the manuscript, or in the decision to submit the manuscript for publication.

Abbreviations

AA	African-American
AIM	Ancestry informative marker
ARTP	Adaptive rank truncated product
BRMS1	Breast cancer metastasis suppressor 1
CD82	CD82 molecule
CDH1	Cadherin 1, type 1, E-cadherin
CEU	Utah residents with Northern and Western European ancestry
CI	Confidence interval
CINJ	Cancer Institute of New Jersey
CTNNB1	Catenin (cadherin-associated protein), beta 1
DNA	Deoxyribonucleic acid
EA	European-American
EMT	Epithelial-mesenchymal transition
ER	Estrogen receptor
ESR1	Estrogen receptor alpha
FDR	False discovery rate
GRASP	Genome-wide repository of associations between SNPs and phenotypes
HER2	Human epidermal growth factor receptor 2

KISS1	KiSS-1 metastasis suppressor
LD	Linkage disequilibrium
LN	lymph node
MAF	Minor allele frequency
MTA1	Metastasis-associated 1
MTA2	Metastasis-associated 1 family, member 2
MTA3	Metastasis-associated 1 family, member 3
NME1	NME/NM23 nucleoside diphosphate kinase 1
OR	Odds ratio
PR	Progesterone receptor
RPCI	Roswell Park Cancer Institute
SATB1	Special AT-rich binding protein homeobox 1
SIPA1	Signal-induced proliferation-associated 1
SNAI1	Snail family zinc finger 1
SNP	Single nucleotide polymorphism
WCHS	Women's Circle of Health Study
YRI	Yoruba in Ibadan, Nigeria

References

1. Higgins MJ, Wolff AC. Therapeutic options in the management of metastatic breast cancer. Oncology (Williston Park). 2008; 22:614–623. discussion 623, 627–9. [PubMed: 18561551]
2. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. CA Cancer J Clin. 2015
3. Daly B, Olopade OI. A perfect storm: How tumor biology, genomics, and health care delivery patterns collide to create a racial survival disparity in breast cancer and proposed interventions for change. CA Cancer J Clin. 2015; 65:221–238. [PubMed: 25960198]
4. Hunter KW, Crawford NP. Germ line polymorphism in metastatic progression. Cancer Res. 2006; 66:1251–1254. [PubMed: 16452174]
5. Nguyen DX, Massagué J. Genetic determinants of cancer metastasis. Nat Rev Genet. 2007; 8:341–352. [PubMed: 17440531]
6. Drasin DJ, Robin TP, Ford HL. Breast cancer epithelial-to-mesenchymal transition: examining the functional consequences of plasticity. Breast Cancer Res. 2011; 13:226. [PubMed: 22078097]
7. Han H-J, Russo J, Kohwi Y, Kohwi-Shigematsu T. SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. Nature. 2008; 452:187–193. [PubMed: 18337816]
8. Bodenstine TM, Welch DR. Metastasis suppressors and the tumor microenvironment. Cancer Microenviron. 2008; 1:1–11. [PubMed: 19308680]
9. Manavathi B, Kumar R. Metastasis tumor antigens, an emerging family of multifaceted master coregulators. J Biol Chem. 2007; 282:1529–1533. [PubMed: 17142453]

10. Tsukamoto N, Hattori M, Yang H, Bos JL, Minato N. Rap1 GTPase-activating protein SPA-1 negatively regulates cell adhesion. J Biol Chem. 1999; 274:18463–18469. [PubMed: 10373454]
11. Park YG, Zhao X, Lesueur F, Lowy DR, Lancaster M, Pharoah P, et al. Sipa1 is a candidate for the metastasis efficiency modifier locus Mtes1. Nat Genet. 2005; 37:1055–1062. [PubMed: 16142231]
12. Ambrosone CB, Ciupak GL, Bandera EV, Jandorf L, Bovbjerg DH, Zirpoli G, et al. Conducting Molecular Epidemiological Research in the Age of HIPAA: A Multi-Institutional Case-Control Study of Breast Cancer in African-American and European-American Women. J Oncol. 2009; 2009:871250. [PubMed: 19865486]
13. Yao S, Zirpoli G, Bovbjerg DH, Jandorf L, Hong C-C, Zhao H, et al. Variants in the vitamin D pathway, serum levels of vitamin D, and estrogen receptor negative breast cancer among African American women. Breast Cancer Res. 2012; 14:R58. [PubMed: 22480149]
14. Bandera EV, Chandran U, Zirpoli G, McCann SE, Ciupak G, Ambrosone CB. Rethinking sources of representative controls for the conduct of case-control studies in minority populations. BMC Med Res Methodol. 2013; 13:71. [PubMed: 23721229]
15. Xu Z, Taylor Ja. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res. 2009; 37:W600–W605. [PubMed: 19417063]
16. Ruiz-Narváez EA, Rosenberg L, Wise LA, Reich D, Palmer JR. Validation of a small set of ancestral informative markers for control of population admixture in African Americans. Am J Epidemiol. 2011; 173:587–592. [PubMed: 21262910]
17. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945–959. [PubMed: 10835412]
18. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995; 57:289–200.
19. Yu K, Li Q, Bergen AW, Pfeiffer RM, Rosenberg PS, Caporaso N, et al. Pathway analysis by adaptive combination of P-values. Genet Epidemiol. 2009; 33:700–709. [PubMed: 19333968]
20. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010; 38:W214–W220. [PubMed: 20576703]
21. Slattery ML, Lundgreen A, Stern MC, Hines L, Wolff RK, Giuliano AR, et al. The influence of genetic ancestry and ethnicity on breast cancer survival associated with genetic variation in the TGF-β-signaling pathway: The Breast Cancer Health Disparities Study. Cancer Causes Control. 2014; 25:293–307. [PubMed: 24337772]
22. Yao Y-L, Yang W-M. The metastasis-associated proteins 1 and 2 form distinct protein complexes with histone deacetylase activity. J Biol Chem. 2003; 278:42560–42568. [PubMed: 12920132]
23. Fu J, Qin L, He T, Qin J, Hong J, Wong J, et al. The TWIST/Mi2/NuRD protein complex and its essential role in cancer metastasis. Cell Res. 2011; 21:275–289. [PubMed: 20714342]
24. Cui Y, Niu A, Pestell R, Kumar R, Curran EM, Liu Y, et al. Metastasis-associated protein 2 is a repressor of estrogen receptor alpha whose overexpression leads to estrogen-independent growth of human breast cancer cells. Mol Endocrinol. 2006; 20:2020–2035. [PubMed: 16645043]
25. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012; 22:1790–1797. [PubMed: 22955989]
26. Dickinson, La, Joh, T., Kohwi, Y., Kohwi-Shigematsu, T. A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. Cell. 1992; 70:631–645. [PubMed: 1505028]
27. Alvarez JD, Yasui DH, Niida H, Joh T, Loh DY, Kohwi-Shigematsu T. The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. Genes Dev. 2000; 14:521–535. [PubMed: 10716941]
28. Heubner M, Kimmig R, Aktas B, Siffert W, Frey UH. The haplotype of three polymorphisms in the SATB1 promoter region impacts survival in breast cancer patients. Oncol Lett. 2014; 7:2007–2012. [PubMed: 24932280]

29. Leslie R, O'Donnell CJ, Johnson AD. GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. Bioinformatics. 2014; 30:i185–i194. [PubMed: 24931982]
30. Shu XO, Long J, Lu W, Li C, Chen WY, Delahanty R, et al. Novel genetic markers of breast cancer survival identified by a genome-wide association study. Cancer Res. 2012; 72:1182–1189. [PubMed: 22232737]
31. Dhasarathy A, Kajita M, Wade PA. The transcription factor snail mediates epithelial to mesenchymal transitions by repression of estrogen receptor-alpha. Mol Endocrinol. 2007; 21:2907–2918. [PubMed: 17761946]
32. Cano A, Pérez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, del Barrio MG, et al. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nat Cell Biol. 2000; 2:76–83. [PubMed: 10655586]
33. Batlle E, Sancho E, Francí C, Domínguez D, Monfar M, Baulida J, et al. The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. Nat Cell Biol. 2000; 2:84–89. [PubMed: 10655587]
34. Yu J-C, Hsu H-M, Chen S-T, Hsu G-C, Huang C-S, Hou M-F, et al. Breast cancer risk associated with genotypic polymorphism of the genes involved in the estrogen-receptor-signaling pathway: a multigenic study on cancer susceptibility. J Biomed Sci. 2006; 13:419–432. [PubMed: 16502042]
35. Andrews JL, Kim AC, Hens JR. The role and function of cadherins in the mammary gland. Breast Cancer Res. 2012; 14:203. [PubMed: 22315958]
36. Lee J-Y, Park AK, Lee K-M, Park SK, Han S, Han W, et al. Candidate gene approach evaluates association between innate immunity genes and breast cancer risk in Korean women. Carcinogenesis. 2009; 30:1528–1531. [PubMed: 19372141]
37. Alanazi MS, Parine NR, Shaik JP, Alabdulkarim HA, Ajaj SA, Khan Z. Association of single nucleotide polymorphisms in Wnt signaling pathway genes with breast cancer in Saudi patients. PLoS One. 2013; 8:e59555. [PubMed: 23516639]
38. Bandyopadhyay S, Zhan R, Chaudhuri A, Watabe M, Pai SK, Hirota S, et al. Interaction of KAI1 on tumor cells with DARC on vascular endothelium leads to metastasis suppression. Nat Med. 2006; 12:933–938. [PubMed: 16862154]
39. Hua K, Feng W, Cao Q, Zhou X, Lu X, Feng Y. Estrogen and progestin regulate metastasis through the PI3K/AKT pathway in human ovarian cancer. Int J Oncol. 2008; 33:959–967. [PubMed: 18949358]
40. Chen S-LL, Wu Y-SS, Shieh H-YY, Yen C-CC, Shen J-JJ, Lin K-HH. P53 is a regulator of the metastasis suppressor gene Nm23-H1. Mol Carcinog. 2003; 36:204–214. [PubMed: 12669312]
41. Curtis CD, Likhite VS, McLeod IX, Yates JR, Nardulli AM. Interaction of the tumor metastasis suppressor nonmetastatic protein 23 homologue H1 and estrogen receptor alpha alters estrogen-responsive gene expression. Cancer Res. 2007; 67:10600–10607. [PubMed: 17975005]
42. Qu S, Long J, Cai Q, Shu X-O, Cai H, Gao Y-T, et al. Genetic polymorphisms of metastasis suppressor gene NME1 and breast cancer survival. Clin Cancer Res. 2008; 14:4787–4793. [PubMed: 18676749]
43. Tan K, Cho S-G, Luo W, Yi T, Wu X, Siwko S, et al. KiSS1-induced GPR54 signaling inhibits breast cancer cell migration and epithelial-mesenchymal transition via protein kinase D1. Curr Mol Med. 2014; 14:652–662. [PubMed: 24894166]
44. Cruz Quevedo EG, Mimendi Aguilar GM, Juárez Aguilar LA, Gutierrez Rubio SA, Flores Martínez SE, Dávalos Rodríguez IP, et al. Polymorphisms rs12998 and rs5780218 in KiSS1 Suppressor Metastasis Gene in Mexican Patients with Breast Cancer. Dis Markers. 2015; 2015:365845. [PubMed: 25810563]
45. Beeghly-Fadiel A, Lu W, Gao Y-T, Long J, Deming SL, Cai Q, et al. E-cadherin polymorphisms and breast cancer susceptibility: a report from the Shanghai Breast Cancer Study. Breast Cancer Res Treat. 2010; 121:445–452. [PubMed: 19834798]
46. Tipirisetti NR, Govatati S, Govatati S, Kandukuri LR, Cingeetham A, Singh L, et al. Association of E-cadherin single-nucleotide polymorphisms with the increased risk of breast cancer: a study in South Indian women. Genet Test Mol Biomarkers. 2013; 17:494–500. [PubMed: 23551055]

47. Lei H, Sjöberg-Margolin S, Salahshor S, Werelius B, Jandáková E, Hemminki K, et al. CDH1 mutations are present in both ductal and lobular breast cancer, but promoter allelic variants show no detectable breast cancer risk. Int J Cancer. 2002; 98:199–204. [PubMed: 11857408]

48. Wang L, Wang G, Lu C, Feng B, Kang J. Contribution of the-160C/A Polymorphism in the E-cadherin Promoter to Cancer Risk: A Meta-Analysis of 47 Case-Control Studies. PLoS One. 2012; 7:e40219. [PubMed: 22792244]

49. Goode EL, Dunning AM, Kuschel B, Healey CS, Day NE, Ponder BaJ, et al. Effect of germ-line genetic variation on breast cancer survival in a population-based study. Cancer Res. 2002; 62:3052–3057. [PubMed: 12036913]

50. Cattaneo F, Venesio T, Molatore S, Russo A, Fiocca R, Frattini M, et al. Functional analysis and case-control study of-160C/A polymorphism in the E-cadherin gene promoter: association with cancer risk. Anticancer Res. 2006; 26:4627–4632. [PubMed: 17201188]

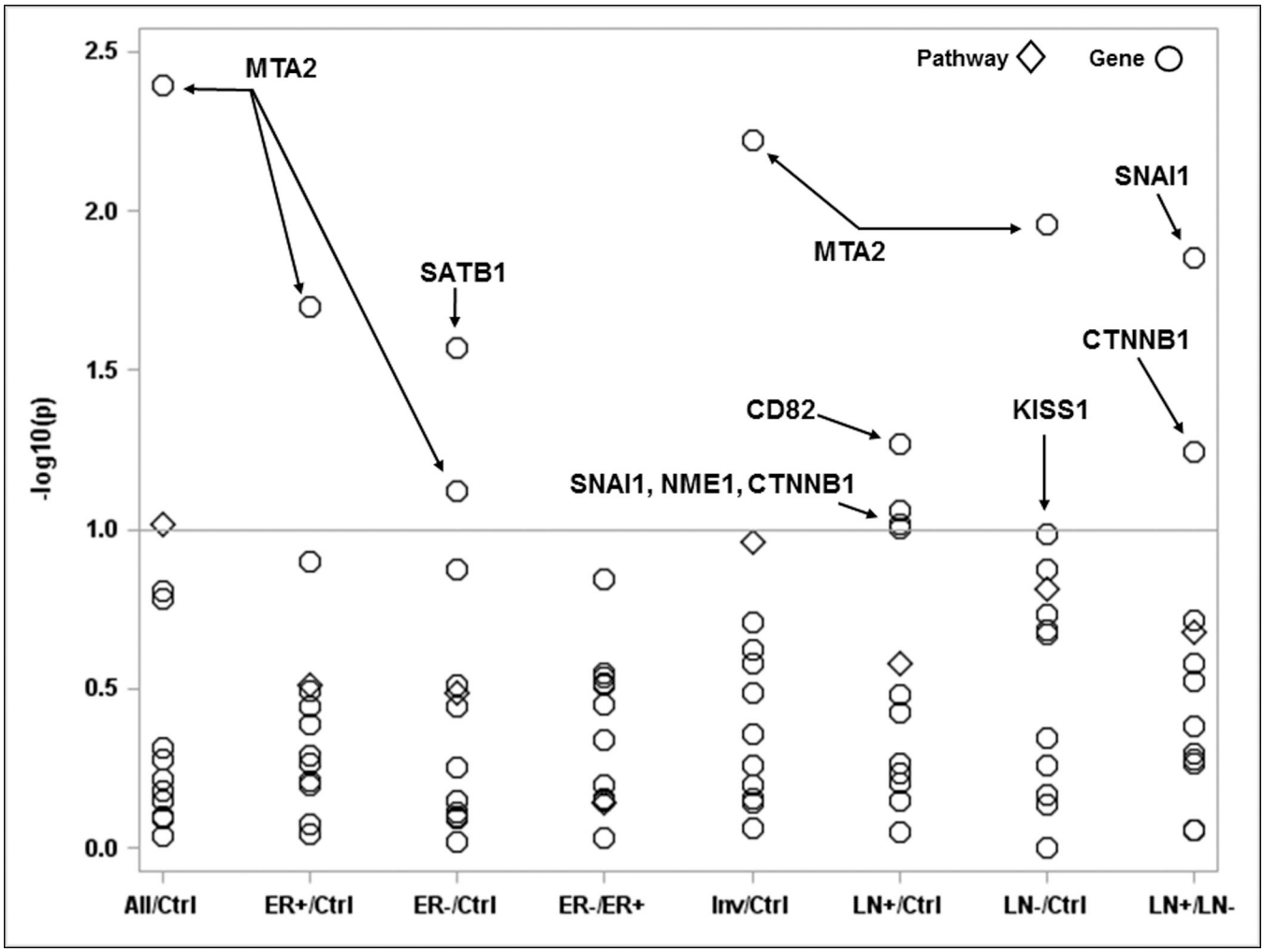
51. Hsieh S-M, Smith Ra, Lintell Na, Hunter KW, Griffiths LR. Polymorphisms of the SIPA1 gene and sporadic breast cancer susceptibility. BMC Cancer. 2009; 9:331. [PubMed: 19765277]

52. Gaudet MM, Milne RL, Cox A, Camp NJ, Goode EL, Humphreys MK, et al. Five polymorphisms and breast cancer risk: results from the Breast Cancer Association Consortium. Cancer Epidemiol Biomarkers Prev. 2009; 18:1610–1616. [PubMed: 19423537]

53. Crawford NPS, Ziogas A, Peel DJ, Hess J, Anton-Culver H, Hunter KW. Germline polymorphisms in SIPA1 are associated with metastasis and other indicators of poor prognosis in breast cancer. Breast Cancer Res. 2006; 8:R16. [PubMed: 16563182]

54. Roberts MR, Hong C-C, Edge SB, Yao S, Bshara W, Higgins MJ, et al. Case-only analyses of the associations between polymorphisms in the metastasis-modifying genes BRMS1 and SIPA1 and breast tumor characteristics, lymph node metastasis, and survival. Breast Cancer Res Treat. 2013; 139:873–885. [PubMed: 23771732]

55. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet. 2008; 9:403–433. [PubMed: 18593304]



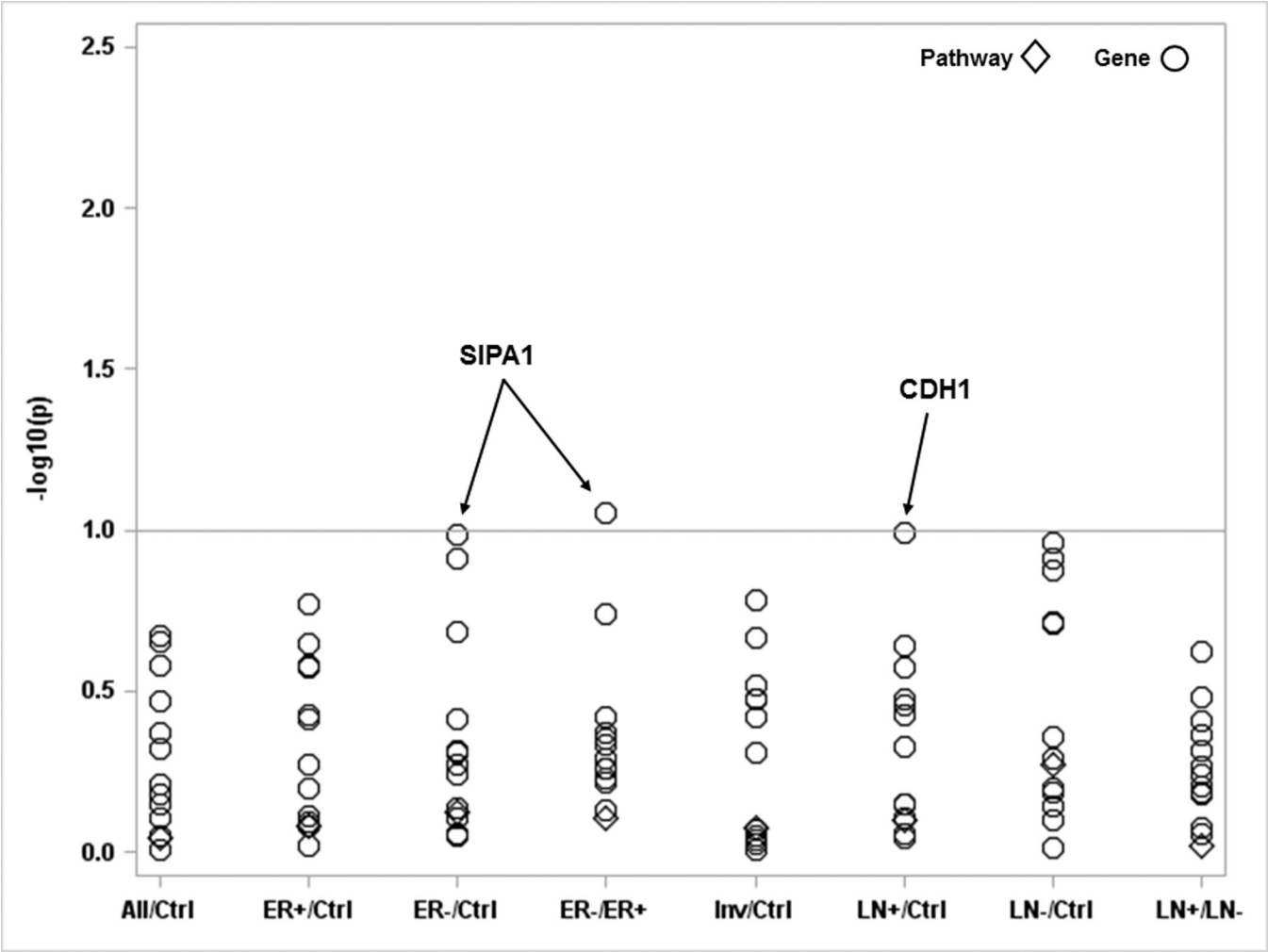


Figure 1. Adaptive rank truncated product method p-values

A. European-American women.

B. African-American women.

Pathway and gene-level permutation p-values were obtained by permuting the response vector 10,000 times for each outcome, using default truncation points. Permutation p-values are shown on the log scale; $-\log(p) \geq 1.0$ (equivalent to $p \leq 0.10$) was considered significant.

Table 1

Selected characteristics of participants in the Women's Circle of Health Study.

Characteristic[a]	EA Controls (n=649)	EA Cases (n=658)	P	AA Controls (n=743)	AA Cases (n=621)	P
Age (y)	49.7 (8.7)	52.1 (10.1)	<0.0001	48.6 (9.4)	51.4 (10.4)	<0.0001
European ancestry (%)	0.985 (0.04)	0.968 (0.08)	<0.0001	0.138 (0.14)	0.137 (0.16)	0.92
Body mass index (kg/m^2)	27.3 (7.2)	27.2 (6.6)	0.63	31.9 (7.8)	31.1 (6.7)	0.03
Education						
High school or less	73 (11.3)	135 (20.5)	<0.0001	295 (39.7)	280 (45.1)	0.18
Some college	117 (18.0)	144 (21.9)		204 (27.5)	167 (26.9)	
College graduate	210 (32.4)	201 (30.6)		149 (20.1)	107 (17.2)	
Post-graduate	249 (38.4)	178 (27.1)		95 (12.8)	67 (10.8)	
Age at first pregnancy[b]						
≤ 19	25 (5.4)	35 (7.8)	0.03	259 (40.8)	195 (37.3)	0.01
20–24	100 (21.6)	118 (26.2)		183 (28.8)	151 (28.9)	
25–30	157 (33.8)	159 (35.3)		99 (15.6)	118 (22.6)	
31 or more	182 (39.2)	139 (30.8)		94 (14.8)	59 (11.3)	
Menopausal status						
Premenopausal	356 (54.9)	343 (52.1)	0.35	411 (55.3)	309 (49.8)	0.04
Postmenopausal	293 (45.2)	315 (47.9)		332 (44.7)	312 (50.2)	
Age at menopause[c]						
≤ 44	25 (8.7)	26 (8.4)	0.39	43 (13.1)	30 (9.9)	0.001
45 – 49	68 (23.6)	59 (19.2)		87 (26.5)	48 (15.8)	
50 or more	195 (67.7)	223 (72.4)		198 (60.4)	226 (74.3)	
History of benign breast disease						
Yes	210 (32.5)	269 (41.3)	0.001	156 (21.0)	193 (31.2)	<0.0001
No	437 (67.5)	382 (58.7)		586 (79.0)	426 (68.8)	
Family history of breast cancer						
Yes	109 (16.8)	161 (24.5)	0.001	87 (11.7)	89 (14.3)	0.17
No	540 (83.2)	497 (75.5)		656 (88.3)	532 (85.7)	

Characteristic[a]	EA Controls (n=649)	EA Cases (n=658)	P	AA Controls (n=743)	AA Cases (n=621)	P
ER status						
Positive	–	387 (82.5)		–	324 (68.4)	
Negative	–	82 (17.5)		–	150 (31.6)	
LN status[d]						
Positive	–	127 (33.2)		–	155 (40.9)	
Negative	–	255 (66.8)		–	224 (59.1)	
Tumor grade						
Low	–	95 (24.4)		–	36 (9.1)	
Intermediate	–	171 (44.0)		–	154 (38.9)	
High	–	123 (31.6)		–	206 (52.0)	

P-values were obtained using Chi-squared tests for categorical variables and t-tests for continuous variables.

[a]Mean (SD) is shown for age, proportion of European ancestry, and body mass index; N (%) is shown for all other variables.

Restricted to

[b]gravid women;

[c]postmenopausal women;

[d]invasive tumors.

Table 2

Multi-allelic risk score associations in European-American and African-American women.

Comparison Gene(s) [N, SNPs]	Case group		Control group		Multivariate model[a]		
	Mean (SD)	Min, Max	Mean (SD)	Min, Max	OR (95% CI)	P	P_{FDR}
European-American							
All cases vs. controls							
MTA2 [2]	1.35 (1.25)	0,4	1.23 (1.23)	0,4	1.15 (1.04–1.26)	0.004	0.01
ER+ cases vs. controls							
MTA2 [2]	1.35 (1.26)	0,4	1.23 (1.23)	0,4	1.16 (1.05–1.30)	0.01	0.01
ER− cases vs. controls							
MTA2/SATB1 [13]	14.40 (3.63)	8,22	13.10 (3.40)	7,23	1.12 (1.05–1.20)	0.001	0.003
Invasive tumor cases vs. controls							
MTA2 [2]	1.39 (1.26)	0,4	1.23 (1.23)	0,4	1.19 (1.07–1.33)	0.001	0.003
LN+ cases vs. controls							
SNAI1, CD82, NME1, CTNNB1 [27]	28.94 (4.50)	19,40	26.90 (4.64)	14,41	1.09 (1.04–1.14)	0.0002	0.001
LN− cases vs. controls							
MTA2, KISS1 [5]	3.59 (1.65)	0,8	3.19 (1.70)	0,10	1.18 (1.08–1.29)	0.0004	0.002
LN+ vs. LN− cases							
SNAI1, CTNNB1 [6]	8.90 (2.13)	2,12	8.30 (2.03)	3,12	1.15 (1.01–1.31)	0.04	0.04
African-American							
ER− cases vs. controls							
SIPA1 [5]	5.81 (1.62)	0,10	5.41 (1.66)	0,9	1.16 (1.02–1.31)	0.02	0.03
ER− vs. ER+ cases							
SIPA1 [5]	6.31 (1.92)	2,10	5.84 (1.79)	2,10	1.23 (0.99–1.52)	0.06	0.06
LN+ cases vs. controls							
CDH1 [19]	22.45 (3.48)	14,30	21.28 (3.57)	10,32	1.13 (1.07–1.19)	<0.0001	0.0003

[a]Adjusted for age, ancestry, and race-specific covariates [EA case-control: state, education, age at first pregnancy, breastfeeding, history of benign breast disease, family history of breast cancer, oral contraceptive use, and body mass index; EA case-case: HER2 status, tumor grade, and tumor size; AA case-control: education, history of benign breast disease, oral contraceptive use, smoking status, hormone replacement therapy use, and age at menopause; AA case-case: PR status, tumor grade, tumor size, and stage].

Multi-allelic scores were constructed for each outcome that had one or more ARTP-significant genes; only SNPs in significant genes were included. Unconditional logistic regression was used to test associations between each score and its corresponding outcome, with FDR p-value correction. SNPs were excluded from score construction if the additive model OR was between 0.95 and 1.05. For AA

women, CDH1 rs16958383 was excluded. For EA women, SNAI1 rs1543442, CD82 rs934178, rs1981994, rs6485550, rs3758813, and rs3781753, KISS1 rs4951315 and rs12568353, and SATB1 rs9852969, rs9849722, rs12106735, and rs6550461 were excluded.

ER=estrogen receptor; FDR=false discovery rate; LN=lymph node; OR=odds ratio; CI=confidence interval.